

## Functional potential of bacterial strains in the premature infant gut microbiome is associated with gestational age

Sumayah F. Rahman<sup>a</sup>, Matthew R. Olm<sup>a</sup>, Michael J. Morowitz<sup>b</sup>, Jillian F. Banfield<sup>c\*</sup>

<sup>a</sup>Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California, USA

<sup>b</sup>Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

<sup>c</sup>Department of Earth and Planetary Sciences and Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, California, USA

\*Corresponding author

E-mail: [jbanfield@berkeley.edu](mailto:jbanfield@berkeley.edu)

### Abstract

The gut microbiota of premature and full-term infants have many known differences, but the extent to which the degree of prematurity influences the structure and functional potential of the microbiome has not been deeply explored. Here, we used genome-resolved metagenomics to address how gestational age impacts the premature infant gut microbiome. We found that gestational age is associated with species richness, with more premature infants having lower species richness; this effect lasts until the fourth week of life. Novel *Clostridium* species and strains related to *Streptococcus salivarius* and *Enterococcus faecalis* colonize infants of different gestational ages, and the metabolic potential of these organisms can be distinguished. Thus, we conclude that the extent of prematurity, or directly linked factors, can be an important influence on the microbiome and its functions.

### Introduction

The human gut microbiome plays many important roles, including the extraction of nutrients from food, metabolizing toxins, immunomodulation, and protection from pathogens<sup>1</sup>. Infants, near-sterile when born, obtain microbes from their mother and their environment<sup>2-4</sup>. The gut microbiome of infants is known for its simplicity and low complexity compared to the gut microbiome of older children and adults<sup>5</sup>. Premature infants, born before they have reached 37 weeks *in utero*, harbor gut microbial communities of even lower complexity than full-term infants, as they are colonized by tenfold fewer species<sup>6</sup>. Premature infant gut microbiomes display abrupt shifts in composition<sup>7</sup>, and may have a different taxonomic makeup than microbiomes of full-term infants<sup>6,8,9</sup>. Studies on premature infants suggest that the extent of the infant's prematurity influences their gut microbiome. A recent study, utilizing 16S rRNA gene sequencing, revealed that bacterial alpha diversity varies based on the infants' gestational age, with more premature infants having less diverse microbiomes. This study also showed that infants born at a later gestational age had greater abundance of *Bifidobacterium* and *Streptococcus*<sup>10</sup>. However, there were no analyses of differences in metabolic potential of microorganisms colonizing infants of different gestational age, due to the low resolution of the 16S method. A metaproteomics study revealed that gut bacteria of extremely preterm infants (< 28 weeks gestational age) produced more translation and membrane transport proteins, while the microbiomes of infants with a gestational age of 30 weeks produced more energy metabolism proteins<sup>11</sup>.

Genome-resolved metagenomics involves sequencing all the DNA extracted from a sample and then reconstructing genomes for the relatively abundant microorganisms present. Previous genome-resolved metagenomics studies found that the infant gut microbiome is influenced by factors such as formula feeding, the hospital room environment, and antibiotic exposure<sup>4,12,13</sup>. Here, we utilize genome-resolved metagenomics to analyze the effects of gestational age on the composition and metabolic potential of the premature infant gut microbiome. It is well-established that prematurity is associated with increased disease and infant mortality<sup>14</sup>, and this is partially due to factors involving the microbiome<sup>15</sup>. Understanding the effect that gestational age, i.e. the extent of prematurity, has on the microbiome may

improve understanding of disease in premature infants. We found that certain bacteria occurring in infants of different gestational ages carry distinct sets of metabolic genes, thus utilizing genome-centric metagenomics to resolve how the gut microbiome is influenced by extent of prematurity.

## Results

We analyzed 900 previously reported samples from 106 premature infants with gestational ages of 24 to 32 weeks at birth (**Table S1**). Samples were collected over the first two to three months of life. Among these infants, just 10% were classified as moderate preterm (defined as 32 to < 34-week gestation), 60% of the infants were very preterm (28 to < 32-week gestation), and 30% were extremely preterm (< 28-week gestation) (**Fig 1a**). A correlation analysis revealed that gestational age is closely associated with birth weight ( $r = 0.84$ ) (**Fig 1b**), but gestational age does not display a correlation with any of the other variables in the infant metadata (**Fig 1c**). Reconstructing genome bins from the 900 samples sequenced resulted in a dereplicated set of 1,483 genomes with an average completeness of 92% as evaluated based on the presence of bacterial single copy genes. There was no significant difference in genome completeness or sequencing depth of infants of different gestational age (**Table S2** and **Table S3**).

A linear regression model was applied to evaluate the effect of the infant's characteristics as well as environmental factors on the species richness of the gut microbial community. As expected, administration of antibiotics due to a disease diagnosis after the first week of life caused a significant decrease in richness ( $p < 1 \times 10^{-6}$ , multiple linear regression analysis) (**Table 1**). The model also revealed that gestational age has a significant effect on species richness ( $p < 1 \times 10^{-6}$ , multiple linear regression analysis) (**Table 1**). To understand how gestational age's effect on species richness changes over the course of the first few months of life, the richness of microbiomes of infants with gestational age < 28 weeks (extremely premature) was compared to that of infants with gestational age  $\geq 28$  weeks, at each week of life. In the first few weeks of life, microbiomes of extremely premature infants have significantly lower richness levels (**Fig 2**). This effect is no longer present at the fifth week of life and onward.

We compared the average taxonomic composition of microbiomes of infants with gestational age < 28 weeks and infants with gestational age  $\geq$  28 weeks (**Fig 3**). The figure shows small fluctuations in composition over time, but the microbiomes of individual infants can display more drastic shifts in composition. In infants with gestational age < 28 weeks, *Klebsiella* was consistently the most abundant taxa, except for during the seventh week of life where *Escherichia* was most abundant (**Fig 3a**). In infants with gestational age  $\geq$  28 weeks, *Escherichia* and *Klebsiella* were initially abundant but the relative abundance of these taxa appeared to decline over time (**Fig 3b**); however, this decline was not statistically significant, as the interindividual variation was substantial. When making direct comparisons of relative abundance between the two infant cohorts in the same week of life, the infants with gestational age  $\geq$  28 weeks had significantly higher populations of *Veillonella*, *Clostridioides* and *Clostridium* throughout the first month of life ( $p < 0.001$ , Mann-Whitney U test with FDR correction). When making comparisons based on corrected age, which is calculated from the time of conception to adjust for prematurity, we found no significant difference between the relative abundance values of *Veillonella* and *Clostridioides* in the two infant cohorts. The lack of difference when matching samples based on corrected age supports the hypothesis that the previously mentioned finding is due to prematurity—as the extremely premature infants reach the second month of life, they become more developmentally similar to the less premature infants, and the (likely prematurity-induced) effects that were present in early life are no longer detectable.

We investigated the differences in metabolic potential of bacterial strains colonizing infants of varying gestational ages. Separate analyses were performed for each week of life and for each species or species group, in the case of *Clostridium*. Within each week we considered only one sample per infant to reduce bias due to resampling of the same strain in subsequent samples, allowing us to test for patterns that were consistent across infants. We uncovered trends related to the extent of prematurity, in which particular organisms exclusively colonized infants of a certain gestational age. In the second week of life, species of a novel group in *Clostridium*, all of which have genes for vitamin B biosynthesis, were only present in infants of gestational age greater than 30 weeks (**Fig 4a**). *Streptococcus salivarius*-related strains containing genes for L-Cystine transport only occurs in infants of gestational age  $\leq$  30 weeks (**Fig 4b**). In the

third and fourth week of life, very-closely related *Enterococcus faecalis* strains with genes for the RaxABRaxC type I secretion system are present exclusively in infants of gestational age  $\geq 28$  weeks (**Fig 4c**). Most of the *E. faecalis* carrying genes for this secretion system also harbor biosynthetic gene clusters for bacteriocin and lantibiotic, while none of the *E. faecalis* lacking the secretory genes were found to harbor these biosynthetic gene clusters.

## Discussion

Our study involved analysis of the microbiomes of 107 premature infants for which a variety of metadata was collected, including each infant's gestational age, birthweight, disease incidence, antibiotic exposure, feeding method, gender, and birth mode (**Table S1**). Agreeing with population-based references built from historical data<sup>16</sup>, the gestational age of infants in this study was closely correlated with birthweight (**Fig 1b**). However, the other collected metadata did not display a correlation with gestational age (**Fig 1c**), indicating that the findings of this microbiome study can be attributed to either (1) the gestational age, directly, or (2) differing clinical treatment among babies of varying gestational age that was not recorded in the collected metadata. This differing clinical treatment could be a particular feeding regimen (e.g., more specific than whether the infant received breastmilk, formula, or a combination), usage of a mechanical ventilator, length and timing of attachment to intravenous nutrition lines, or another factor.

Regardless of whether it is a direct or indirect effect, gestational age was strongly associated with species richness (**Table 1**), and this effect is only present during the first month of life (**Fig 2**). This indicates that while gestational age influences the microbes present in the few weeks immediately following birth, it does not have a persistent impact on microbiome complexity over the study period. We found that no other factors besides gestational age and post-week infant antibiotic exposure had a significant impact on richness of the gut microbiome (**Table 1**). This contrasts with findings of previous studies that associated formula feeding with increased diversity<sup>17</sup> and intrapartum maternal antibiotic use with decreased diversity<sup>18</sup>. It is important to note, however, that in the current study, organisms can be grouped at the strain or

species level, whereas prior studies mostly relied on 16S rRNA gene fragment profiling, which typically has genus-level resolution.

Perhaps the most surprising result was that the day of life (infant's age in days) did not have a significant influence on species richness (**Table 1**). Previous studies have shown that the microbiome of full term infants gains species over time and displays a clear increase in diversity<sup>19</sup>. The difference between the patterns reported here and prior studies may be largely accounted for by prematurity and, in some cases the administration of antibiotics that cause a sharp decline in microbiome diversity<sup>20</sup>. Because these infants were in the neonatal intensive care unit throughout our study, the consortia available to colonize them likely had lower diversity than would have been encountered in the home environment and includes bacteria considered to be hospital-associated pathogens<sup>21</sup>.

We found that certain organisms only colonized infants of a particular gestational age range. One such case is *Clostridium*, which are anaerobes that have been previously found in the infant gut<sup>22</sup>. We identified a group of *Clostridium* that was not closely related to previously sequenced organisms, and found that some organisms in this group harbor genes for biosynthesis of pantothenate, also called vitamin B<sub>5</sub>, a water-soluble vitamin and an essential nutrient typically supplied by intestinal bacteria<sup>23</sup>. These novel *Clostridium* species with genes for pantothenate production were only found in less premature infants (**Fig 4a**). It should be noted that two of the infants harboring *Clostridium* with the pantothenate biosynthesis genes are twins, born at a gestational age of 32 weeks. Genetic relatedness, exposure to the same mother's microbiota, or other factors may have contributed to colonization by similar bacteria. However, the other infants showing gestational age-dependent colonization were unrelated. The observation that very premature infants may have comparatively lower access to vitamin B<sub>5</sub> than less premature infants due to strain colonization may be important because lack of pantothenic acid can adversely affect the immune system, producing a pro-inflammatory state<sup>24</sup>. It has been shown previously that the production of pantothenate in the gut is negatively impacted by low availability of vitamin D, which is often the case with very premature infants<sup>25</sup>. Thus, selection against *Clostridium* strains with the capacity for pantothenate production may be explained by increased prematurity.

*Streptococcus salivarius*-related bacteria with genes for transport of L-cystine, an amino acid essential for infants, are present only in infants of less than 31 weeks gestational age (**Fig 4b**). The infants in this study received cysteine, which forms the cystine dimer, as part of an amino acid mixture included in the parenteral nutrition. A study evaluating plasma amino acid concentrations in infants given parenteral nutrition found that infants of lower birthweight have less of an ability to use cystine/cysteine compared to infants of higher birthweight<sup>26</sup>. Since the more premature infants have lower birthweight (**Fig 1b**), the inability of the human cells to uptake cystine may lead to higher concentrations of cystine in the gut, indicating why cystine-transporting bacteria are selected for in these infants.

The trends described with *Clostridium* and *S. salivarius* both occur in the second week of life. However, *Enterococcus faecalis* displays interesting pattern in the third and fourth week of life: strains with a RaxABRaxC type I secretion system occur only in infants of greater gestational age, while strains lacking this secretion system occur exclusively in extremely premature infants (**Fig 4c**). The RaxABRaxC type I system is involved in the secretion of double-glycine-type leader peptides<sup>27</sup>, which occur in bacteriocins and lantibiotics<sup>28</sup>. All the *E. faecalis* with the RaxABRaxC type I secretion system have gene clusters for production of lantipeptides, bacteriocin, or both. In contrast, most of the *E. faecalis* lacking the secretion system (i.e., the *E. faecalis* occurring in the infants of lowest gestational age) did not have these biosynthetic gene clusters. Since bacteriocins are toxins that inhibit the growth of closely related strains, the gestational age of an infant could indirectly influence the contribution of *E. faecalis* to bacteriocin production and thus influence microbiome composition. As *E. faecalis* is a common and often abundant member of the gut microbiomes of premature infants<sup>29</sup>, it is possible that bacteriocin production is less common in infants of very low gestational ages.

The findings discussed above offer a strain-level perspective to what is known about how the taxonomic and functional characteristics of the gut microbial community change depending on the infant's gestational age<sup>10,11</sup>. By analyzing the metabolic potential of each genome, we found evidence that the extent of prematurity, either directly or indirectly, can affect the gut microbiome. Given evidence that a lower gestational age may limit bacteriocin and vitamin production, which are factors that can impact community structure and lead to inflammation<sup>30,31</sup>,

these findings may inform our understanding of diseases associated with dysbiosis of the microbiome, especially in very premature infants.

## Methods

Sample collection and metagenomic data processing for these samples were previously described<sup>32</sup>. Briefly, fecal samples were collected from premature infants residing in the neonatal intensive care units (NICU) of the Magee-Women's Hospital in Pittsburgh, PA, and a PowerSoil DNA isolation kit (Mo Bio Laboratories, Carlsbad, CA) was used to extract the DNA, which was then sequenced on an Illumina platform (further details available here<sup>21,33,34</sup>). The samples analyzed in this study have been previously reported, and the reads are publicly available at NCBI as described here<sup>12,21,33,34</sup>.

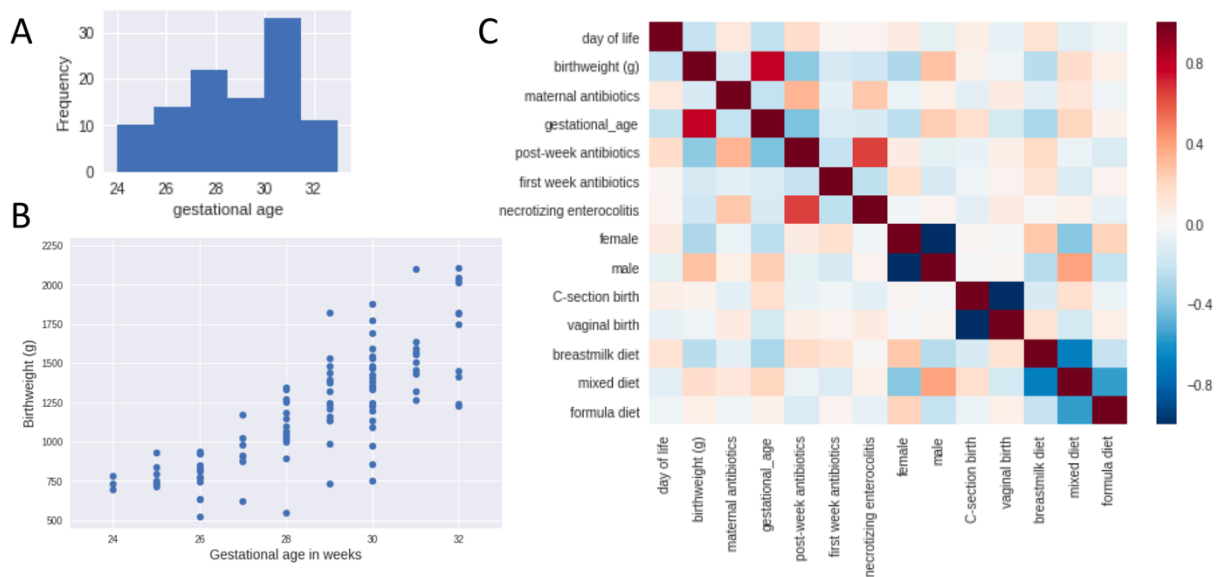
The reads were trimmed using Sickle (<https://github.com/najoshi/sickle>) and cleared of human contamination through read mapping with Bowtie2<sup>35</sup>. IDBA-UD<sup>36</sup> was used to assemble the reads of each sample and was also used to generate co-assemblies by combining the reads of all the samples from a particular infant. The genes on the scaffolds were predicted using Prodigal<sup>37</sup>. The scaffolds were grouped into genome bins using concoct<sup>38</sup> and redundant bins were dereplicated using dRep<sup>39</sup> v0.4.0. Centrifuge<sup>40</sup> was used for initial assignment of taxonomy, which was then verified by confirming the phylogenetic profiles of the scaffolds in the genome bin. When majority of the scaffolds in a genome bin were of unknown phylogenetic profile, the organism was concluded to be novel. For annotation of metabolic capabilities, the sequences were searched against the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>41</sup> using profile hidden Markov models, and the results were used to generate a KEGG metabolism profile for each organism that displayed the fraction of each KEGG module encoded by that genome. For specialized identification of biosynthetic gene clusters, antiSMASH<sup>42</sup> was used to annotate genes from particular scaffolds of interest. Organisms were considered to be present in a particular sample if the genome bin showed breadth of coverage over 99%.

When calculating correlations, Pearson's product-moment correlation was used for two continuous variables and the point biserial correlation was used for pairs that contained at least

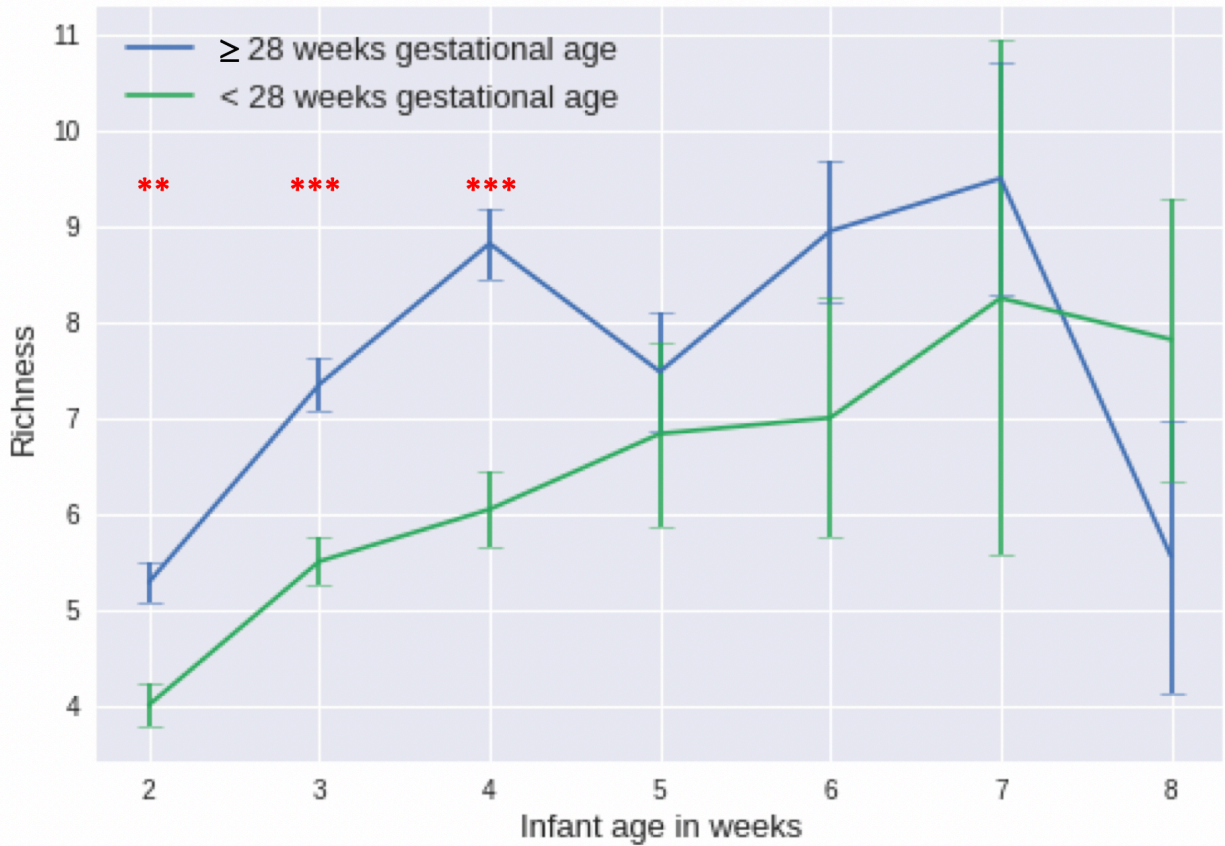


one categorical variable. To model the effects of infant characteristics and clinical treatments on gut species richness levels, linear regression from the scikit-learn<sup>43</sup> package was utilized. Mann-Whitney U tests were performed for comparison of relative abundance values and for the comparison of richness values at each week, and p-values were corrected using the False Discovery Rate (FDR) method. Spearman correlations were performed to identify trends over time. When carrying out the strain-focused portion of the study, separate analyses were performed for each week of life and for each species, to remove effects of these factors. Only one sample per infant was included in each analysis to avoid bias due to repeated samples. When a genome was considered to be lacking a particular gene, the entire dRep set (all genomes belonging to the same secondary cluster in dRep) was checked to ensure that another genome in that set did not harbor the gene of interest. Associations were considered spurious and removed if the validation step did not confirm the findings.

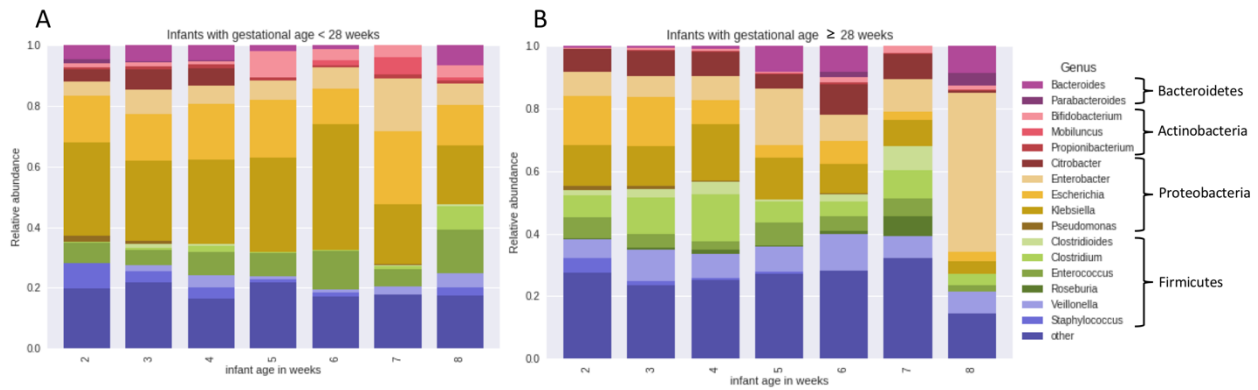
## Figures



**Figure 1.** Gestational age and related infant metadata. (A) The distribution of gestational age in weeks. (B) Gestational age is correlated with birthweight ( $r = 0.84$ ,  $p < 1 \times 10^{-26}$ , Pearson correlation). (C) Gestational age is not strongly correlated with any variable other than birthweight.



**Figure 2.** Mann-Whitney U tests were performed to compare species richness in infants with < 28 week gestational age and infants  $\geq$  28 week gestational age, at each week of life. Two asterisks indicate statistical significance at  $p < 0.005$ , while 3 asterisks indicate statistical significance at  $p < 0.0005$ . Error bars represent standard error of the mean. The number of samples in each week for infants < 28 weeks are as follows: week 2  $n = 79$ , week 3  $n = 91$ , week 4  $n = 79$ , week 5  $n = 13$ , week 6  $n = 14$ , week 7  $n = 5$ , week 8  $n = 12$ . The number of samples in each week for infants  $\geq$  28 weeks are as follows: week 2  $n = 190$ , week 3  $n = 170$ , week 4  $n = 85$ , week 5  $n = 29$ , week 6  $n = 22$ , week 7  $n = 10$ , week 8  $n = 10$ .



**Figure 3.** (A) The genus-level taxonomic composition of the gut community for the infants with < 28 week gestational age. The number of samples in each week for infants < 28 weeks are as follows: week 2  $n = 79$ , week 3  $n = 91$ , week 4  $n = 79$ , week 5  $n = 13$ , week 6  $n = 14$ , week 7  $n = 5$ , week 8  $n = 12$ . (B) The genus-level taxonomic composition of the gut community for the infants with  $\geq 28$  week gestational age. The number of samples in each week for infants  $\geq 28$  weeks are as follows: week 2  $n = 190$ , week 3  $n = 170$ , week 4  $n = 85$ , week 5  $n = 29$ , week 6  $n = 22$ , week 7  $n = 10$ , week 8  $n = 10$ .



**Figure 4.** Trends in strain functional potential related to extent of prematurity. Each dot represents one genome, and the location of the dot on the y-axis indicates the gestational age of the infant in which the organism was found. Green dots indicate that the genome has the particular metabolic function listed in the title of the plot, and blue dots indicate that the genome lacks this function. (A) Genomes part of a novel group in *Clostridium* labeled as having or lacking genes for pantothenate biosynthesis. (B) *Streptococcus salivarius*-related strains labeled as having or lacking genes for an L-cystine transport system. (B) Strains of *Enterococcus faecalis* labeled as having or lacking genes for RaxAB-RaxC type 1 secretion system.

	<b>Coefficient Estimate</b>	<b>p-value</b>
intercept	-4.819817	0.999999
infant age (days)	0.096755	0.558479
gestational age (weeks)	0.385138	0.000000
birthweight (g)	-0.001211	0.979418
maternal antibiotics	-0.321744	0.215535
antibiotic exposure during first week	0.005125	0.915728
post-week antibiotic exposure	-1.427714	0.000000
male gender	0.098894	1
vaginal birth	0.141521	1
breastmilk	0.566750	1
formula	-1.014005	1
combination of breastmilk and formula	0.447255	1

**Table 1:** Linear regression using infant metadata variables to predict species richness.

## Author Contributions

JFB and MJM conceived of the project. MRO processed the sequence data. SFR performed the data analysis and interpreted the results. SFR and JFB wrote the manuscript, and all authors contributed to manuscript revisions.

## Acknowledgements

We acknowledge Robyn Baker for recruiting infants, Brian Firek for performing DNA extractions, Christopher Brown for scripts to calculate genome coverage, and David Burstein for the KEGG HMM annotation pipeline.

## Funding Statement

Funding was provided through the National Institutes of Health (NIH) under grant RAI092531A and the Alfred P. Sloan Foundation under grant APSF-2012-10-05. This work used the Vincent J. Coates Genomics Sequencing Laboratory, supported by NIH S10 OD018174 Instrumentation Grant. The study was approved by the University of Pittsburgh Institutional Review Board (IRB) (Protocol PRO12100487).

## Literature Cited

1. Jandhyala, S. M. *et al.* Role of the normal gut microbiota. *World J. Gastroenterol.* **21**, 8836–8847 (2015).
2. Makino, H. *et al.* Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota. *PLoS One* **8**, (2013).
3. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci.* **108**, 4578–4585 (2011).
4. Brooks, B. *et al.* Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome* **2**, 1 (2014).
5. Yatsunencko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature*

- 486**, 222–228 (2012).
6. Gibson, M. K. *et al.* Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 16024 (2016).
  7. Costello, E. K., Carlisle, E. M., Bik, E. M., Morowitz, M. J. & Relman, D. A. Microbiome Assembly across Multiple Body Sites in Low-Birthweight Infants. *MBio* **4**, (2013).
  8. Rodríguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Heal. Dis.* **26**, 1–17 (2015).
  9. Sim, K. *et al.* The neonatal gastrointestinal microbiota: the foundation of future health? *Arch. Dis. Child. - Fetal Neonatal Ed.* **98**, F362 LP-F364 (2013).
  10. Chernikova, D. A. *et al.* The premature infant gut microbiome during the first 6 weeks of life differs based on gestational maturity at birth. *Pediatr. Res.* **84**, 71–79 (2018).
  11. Zwiittink, R. D. *et al.* Metaproteomics reveals functional differences in intestinal microbiota development of preterm infants. *Mol. Cell. Proteomics* **16**, 1610–1620 (2017).
  12. Rahman, S. F., Olm, M. R., Morowitz, M. J. & Banfield, J. F. Machine learning leveraging genomes from metagenomes identifies influential antibiotic resistance genes in the infant gut microbiome. *mSystems* **3**, (2018).
  13. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).
  14. Kramer, M. S. *et al.* The contribution of mild and moderate preterm birth to infant mortality. Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. *Jama* **284**, 843–9 (2000).
  15. Morrow, A. L. *et al.* Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome* **1**, 13 (2013).
  16. Talge, N. M., Mudd, L. M., Sikorskii, A. & Basso, O. United States Birth Weight Reference Corrected For Implausible Gestational Age Estimates. *Pediatrics* **133**, 844–853 (2014).
  17. Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z. & Maria, G. The infant microbiome development: mom matters. *Trends Mol Med* **21**, 109–117 (2015).
  18. Mshvildadze, M., and Neu, J. The Infant Intestinal Microbiome: Friend or Foe? *Early Hum. Dev.* **86**, 67–71 (2010).

19. Bäckhed, F. *et al.* Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life (*Cell Host and Microbe* (2015) 17(5) (690–703)). *Cell Host Microbe* **17**, 852 (2015).
20. Relman, D. A. The human microbiome: ecosystem resilience and health. *Nutr Rev.* **70**, 1–12 (2012).
21. Brooks, B. *et al.* Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat. Commun.* **8**, 1–7 (2017).
22. Ferraris, L. *et al.* Clostridia in premature neonates' gut: Incidence, antibiotic susceptibility, and perinatal determinants influencing colonization. *PLoS One* **7**, (2012).
23. Said, H. M. Intestinal absorption of water-soluble vitamins in health and disease. *Biochem. J.* **437**, 357 LP-372 (2011).
24. Gominak, S. C. Vitamin D deficiency changes the intestinal microbiome reducing B vitamin production in the gut. The resulting lack of pantothenic acid adversely affects the immune system, producing a “pro-inflammatory” state associated with atherosclerosis and autoimmun. *Med. Hypotheses* **94**, 103–107 (2016).
25. Gurmeet, S. Vitamin D levels in preterm and term neonates at birth. *Int. J. Contemp. Pediatr.* **4**, 48–52 (2017).
26. W.C., H. *et al.* Pediatric parenteral amino acid mixture in low birth weight infants. *Pediatrics* **81**, 41–50 (1988).
27. da Silva, F. G. *et al.* Bacterial Genes Involved in Type I Secretion and Sulfation Are Required to Elicit the Rice *Xa21* -Mediated Innate Immune Response. *Mol. Plant-Microbe Interact.* **17**, 593–601 (2004).
28. Aymerich, T. *et al.* Biochemical and genetic characterization of enterocin A from *Enterococcus faecium*, a new antilisterial bacteriocin in the pediocin family of bacteriocins. *Appl. Environ. Microbiol.* **62**, 1676–1682 (1996).
29. Moles, L. *et al.* Preterm infant gut colonization in the neonatal ICU and complete restoration 2 years later. *Clin. Microbiol. Infect.* **21**, 936.e1-936.e10 (2015).
30. Umu, Ö. C. O. *et al.* The potential of class II bacteriocins to modify gut microbiota to improve host health. *PLoS One* **11**, 1–22 (2016).



31. Hibberd, M. C. *et al.* The effects of micronutrient deficiencies on bacterial species from the human gut microbiota. *Sci. Transl. Med.* **9**, (2017).
32. Rahman, S. F., Olm, M. R., Morowitz, M. J. & Banfield, J. F. Machine learning leveraging genomes from metagenomes identifies influential antibiotic resistance genes in the infant gut microbiome. *bioRxiv* (2017).
33. Raveh-Sadka, T. *et al.* Evidence for persistent and shared bacterial strains against a background of largely unique gut colonization in hospitalized premature infants. *ISME J.* **10**, 2817–2830 (2016).
34. Raveh-Sadka, T. *et al.* Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *Elife* **2015**, e05477 (2015).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
36. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
37. Hyatt, D. *et al.* Prodigal : prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
38. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1154 (2014).
39. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: A tool for fast and accurate genome de-replication that enables tracking of microbial genotypes and improved genome recovery from metagenomes. *ISME J.* (2017).
40. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Res.* **26**, 1721–1729 (2016).
41. Kanehisa, M. & Goto, S. KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
42. Medema, M. H. *et al.* AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, 339–346 (2011).

43. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012).